Journal of Statistical Physics, Vol. 44, Nos. 1/2, 1986

Rigorous Statistical Procedures for Data from Dynamical Systems

Manfred Denker[1] and Gerhard Keller[2]

Received May 3, 1985; final February 5, 1986

Various questions about the invariant measures of a dynamical system can be answered by computations of regular functionals or by ranking methods based on a set of observations. This includes symmetry tests and the determination of dimension coefficients. The paper contains the theoretical results and several simulations explain the methods.

KEY WORDS: Dynamical systems; invariant measure; Hausdorff dimension; symmetric densities; U statistics; rank statistics.

1. INTRODUCTION

There is a growing common interest in quite different fields of science in dynamical systems whose trajectories show highly complex behavior. Among the earliest, and by now most popular examples, are the Lorenz equations[22] a system of three coupled, nonlinear ordinary differential equations, which was designed as a simplified model for convection (see Sparrow[36] for a detailed discussion). Other classical examples are the population models described by May[24,25] and many further references can be found.[11,30]

A common feature of all these models is that, depending on the choice of some governing parameters, they show very different kinds of asymptotic behavior, ranging from convergence to a single stable state via limit cycle behavior to what is now called "chaos." Although there are several approaches to the concept of chaos, we single out two ideas which are most

[1] Georg-August-Universitaet, Institut fuer Mathematische Stochastik, Lotzestrasse 13, D-3400 Goettingen, Germany.

[2] Institut fuer Angewandte Mathematik, Im Neuenheimer Feld 294, D-6900 Heidelberg, Germany.

often associated with it: (a) there are uncountably many trajectories which are far from being periodic,[21] and (b) there are "many" (in a suitable sense) trajectories which are sensitive to initial conditions, i.e., where arbitrarily small changes in the trajectory's initial value result in a completely different trajectory after some time.[10,21] Although these concepts are of topological nature, many authors claim "randomness" for chaotic systems, and indeed, visual inspection of computer simulations of such systems seems to justify this terminology.[1] A more rigorous description of randomness in deterministic dynamical systems is provided by the ergodic theoretical approach. Besides its topological structure the phase space is endowed with a certain canonical measure (e.g., Lebesgue or Liouville measure), and mixing properties of the dynamics with respect to this measure specify the degree of stochasticity of the system. Existence or nonexistence of an invariant probability measure absolutely continuous with respect to the canonical measure is closely related to this problem. This invariant measure has an invariant density function[3] with respect to the canonical measure describing, by Birkhoff's ergodic theorem, the asymptotic distribution of a typical trajectory in the phase space. In general these investigations are more difficult than those of topological nature. Only relatively simple models have been studied so far (see Refs. 3, 32, 33, and 39) among these certain classes of piecewise monotonic transformations of an interval.[15,16,28,37] For some of these systems, in particular for the interval transformations, a type of probabilistic mixing property has been established that is known as the weak Bernoulli property in ergodic theory and as absolute regularity in probability theory. This fact has an important consequence: A sequence of successive observations made on such a system can be regarded as the outcome of a stationary sequence of random variables. To be more precise, this stationary sequence forms a so-called functional of some absolutely regular stationary process (cf. Ref. 17 for definitions). In this paper, by a "process" we mean a sequence of random variables defined on some probability space, for example, the phase space together with its invariant measure. This describes our general point of view, and Section 2 gives a precise formulation. Since many of the more complicated systems behave (in the above sense) very much like an interval transformation (for the Lorenz equations this is discussed in Ref. 36), there is some evidence to suggest that for many systems the typical trajectories may be interpreted as realizations of a functional of an absolutely regular process. Hence, many of the classical asymptotic results in probability are available for these processes, like the central limit theorem[17] or the invariance principle.[34]

[3] The term density or density function will only be used with this probabilistic meaning.

The aim of this paper is to show that various questions about the invariant measures of dynamical systems can be answered by applying classical statistical methods to the successive outcomes of the systems just as in the case of statistically independent observations. The first steps into this direction were made by the authors several years ago (see Ref. 4 for a discussion), and the theoretical part of the present note is based on Ref. 5. We give two examples:

1. Consider the transformations T_1, T_2 on $[0, 1]$ given by $T_1(x) = 1 - |2x - 1|$, $T_2(x) = 1 - |2x - 1|^2 = 4x(1 - x)$ (see Fig. 1). Both are symmetric (at $\frac{1}{2}$) and have symmetric invariant densities $h_1(x) = 1$ and $h_2(x) = 1/\pi \sqrt{x(1-x)}$. The symmetry of the densities is not a consequence of the symmetry of the transformations. Indeed, Györgyi/Szépfalusy[12] consider perturbations of T_1, T_2 that are symmetric again although their invariant densities are not

$$T_i(\varepsilon, x) = T_i(x) + \varepsilon T_i(x)[1 - T_i(x)] \qquad i = 1, 2, |\varepsilon| < \varepsilon_0$$

The corresponding invariant densities are analytic in ε[19] and have expansions $h_1(\varepsilon, x) = 1 + \varepsilon(2x - 1) + O(\varepsilon^2)$ and $h_2(\varepsilon, x) = h_2(x)[1 + \varepsilon$

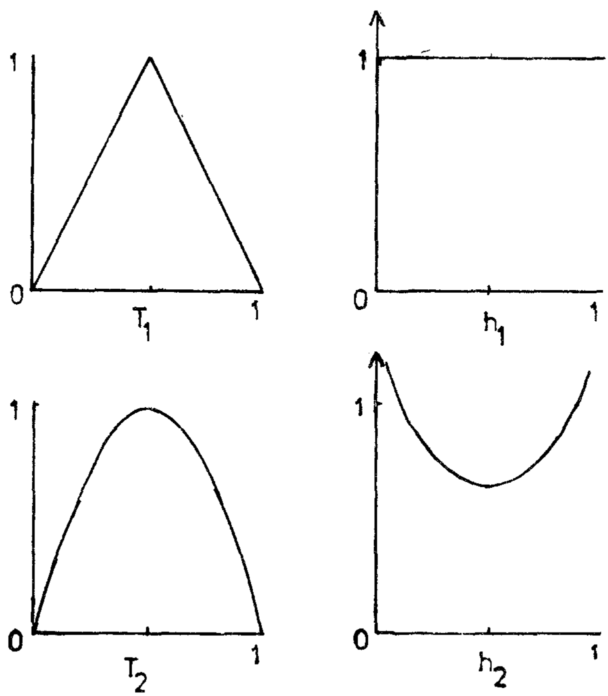

Figure 1

$(x-\frac{1}{2})] + O(\varepsilon^2)$. When such a system is observed, it is of interest whether the density is symmetric ($\varepsilon = 0$) or not ($\varepsilon \neq 0$). In Section 4 we compare two test procedures for this problem based on the sample mean and the Wilcoxon signed rank statistic.

2. Many higher-dimensional systems have an attracting subset of the phase space which is of fractal dimension.[23] Based on numerical simulations and on rigorous results for some special cases, a general relation between this fractal dimension and the Lyapunov numbers of the dynamics is conjectured.[7] This is interesting because the dimension of the attracting set is a geometric constant that is defined without explicit reference to the dynamics of the system, whereas the Lyapunov numbers are dynamical quantities describing an aspect of the time evolution of the system.

In higher-dimensional systems there is no effective way for calculating the fractal dimension from observations.[9] Grassberger and Procaccia (Ref. 8) suggested that a related quantity, they call correlation exponent, and that in many cases agrees with the fractal dimension be studied. In contrast to the dimension it reflects differences in concentration of the invariant distribution living on the attractor of the system: Let X be the phase space, F the invariant distribution. For $r > 0$ define the spatial correlation integral by

$$C(r) = \int_X \int_X 1_{\{\text{dist.}(x,\, y) < r\}}\, dF(x)\, dF(y) \tag{1.1}$$

(dist. denotes the distance in phase space). Note that $C(r)$ measures the concentration of F and describes the mean volume of a ball of radius r. Grassberger and Procaccia[8] show that in many cases $C(r) \simeq \text{const.}\, r^\nu$ $(r \to 0)$ for some $\nu > 0$ and that this ν is called the correlation exponent. For a detailed discussion of dimension see Ref. 6.

In order to determine ν experimentally one has to estimate $C(r)$ for different values of r from observations $X_1,..., X_N$. This is done using the estimator

$$C_N(r) = \frac{1}{N(N-1)} \sum_{\substack{i,j=1\\ i \neq j}}^{N} 1_{\{\text{dist.}(X_i, X_j) < r\}} \tag{1.2}$$

which is in fact a U-statistic estimator.

It can be shown that $C_N(r)$ converges to $C(r)$ for almost all sequences $X_1, X_2,...$ of observations; in fact, convergence in probability follows from the results in this note. Consequently (1.1) agrees with the usual definition. Having estimated $C_N(r)$ for different rs, a linear regression will determine ν.

In particular, it will be shown that error bounds derived from U-statistic theory behave much better than the mean square deviation (which can be applied very doubtfully in this case). Thus, fewer observations are necessary to obtain estimates with reliable error bounds.

In Section 2 we give the general definitions of the properties "weak Bernoulli" and "absolutely regular" and show by a simple example how to regard a dynamical system stochastically as a functional of an absolutely regular process. Our theoretical results are stated in Section 3; in particular, we give explicit conditions that are fulfilled for all the examples in Section 4, where we present results of simulations. We show that the asymptotic theory is applicable for reasonable sample sizes. Our main example is the symmetry problem for invariant densities of interval maps introduced above: We compare the simulation results of two different symmetry tests to their theoretically computed values. We also determine their relative asymptotic efficiency for a broad class of transformations, which measures the power of these test procedures. Then, by means of a simple estimation problem, we demonstrate how seriously the nonvanishing correlations between successive observations may influence the asymptotic variance of the statistic under consideration and, finally, we give some simulation results for the estimator of the correlation exponent from example II (Section 4) using a simple two-dimensional model[18]

$$T(x, y) = (2x \bmod 1, \lambda y + \cos 4\pi x), \qquad \text{where } 0 < \lambda < 1 \tag{1.3}$$

As a final remark note that the y component of $(x_n, y_n) = T^n(x, y)$ may be considered as an autoregressive process of order 1 with error-term $\cos 2\pi x_n$, where (x_n) is a stationary process with good mixing properties.

2. STOCHASTICITY OF DYNAMICAL SYSTEMS

We begin with a very simple model explaining the situation: Consider the transformation $T: x \to 2x \pmod 1$ of the interval $[0, 1)$, and fix an initial value x_0. ($x_n = T^n x_0$ is the observed process.) The map T associates to x_0 a sequence of digits $a_0, a_1, a_2, \ldots$ (its itinerary) by the following rule

$$a_i = 0 \quad \text{if} \quad T^i x_0 \in [0, \tfrac{1}{2}), \qquad a_i = 1 \quad \text{if} \quad T^i x_0 \in [\tfrac{1}{2}, 1)$$

Note that $0.a_0 a_1 a_2 a_3 \cdots$ is the binary expansion of x_0, i.e.

$$x_0 = \sum_{i=0}^{\infty} a_i 2^{-(i+1)} \tag{2.1}$$

Endowed with the Lebesgue measure, $[0, 1)$ is a probability space, and through their dependence on x_0 the a_i become independent identically distributed random variables with $P(a_i=0)=P(a_i=1)=\frac{1}{2}$. From (2.1) we conclude that

$$T^n x_0 = \sum_{i=0}^{\infty} a_{n+i} 2^{-(i+1)}$$

Thus the deterministic sequence $(T^n x_0)_{n\geqslant 0}$ is realized as a functional of the i.i.d. process $(a_i)_{i\geqslant 0}$. The stationarity of $(T^n x_0)_{n\geqslant 0}$ is due to the T invariance of the Lebesgue measure on $[0, 1)$, while the independence of the a_i reflects the good mixing properties of T.

In general the situation is more complicated: Let m be a probability measure on the space X and $T: X\to X$ a measurable transformation (invertible or not) which leaves the measure m invariant, i.e., $m(T^{-1}A)=m(A)$ for $A\subseteq X$. Fix a finite partition $I=(I_1,..., I_N)$ of X into (measurable) sets and associate with each $x_0\in X$ its itinerary (a_n) by the rule "$a_n=i$ if $T^n x_0\in I_i$"; n ranges over the integers if T is invertible and over the nonnegative integers if T is noninvertible. The sequence $(a_n)=(a_n(x_0))$ is a stationary process with respect to the measure m. It is called weak Bernoulli if the coefficients

$$\beta_k = \sup_{s\geqslant 0}\ \sup_{t\geqslant 0} \sum_A \sum_B |m(A\cap B)-m(A)\,m(B)|$$

tend to zero as $k\to\infty$, where the sums are taken over all As of the form $\{x\in X: a_i(x)=c_i,\ 0\leqslant i\leqslant s\}$ and over all Bs of the form $\{x\in X: b_j(x)=d_j,\ s+k\leqslant j\leqslant s+k+t\}$ where the c_i and d_j belong to the set of possible values of the itineraries. Intuitively, β_k measures globally how far apart from independence two pieces of the itineraries are when one of the pieces is observed k units of time after the last observation of the other one. $\beta_k/2$ may be taken as a definition of the coefficient of absolute regularity of the stationary process (a_n).[2,38] Hence (a_n) is a weak Bernoulli process if and only if it is absolutely regular, and the corresponding mixing coefficients coincide up to a factor $\frac{1}{2}$. For certain piecewise expanding interval maps (including those discussed in this paper) it is known that $\beta_k\to 0$ exponentially fast as $k\to\infty$;[15,16] a sketch of the method used there is given in Appendix I.

We still have to make sure that the sequence $(T^n x_0)$ can be recovered from the process (a_n), i.e., that the labeling $X\to\{1,..., N\}^{\mathbb{N}\,(\text{or}\,\mathbb{Z})}$ defined by the itineraries is injective (not necessarily onto). For piecewise expanding interval maps this is easily checked (cf. Theorem 5, Ref. 15). Hence, for these transformations there is some (partially defined) $\Phi: \{1,..., N\}^{\mathbb{N}}\to X$ such that $T^n x_0=\Phi(a_n, a_{n+1}, a_{n+2},...)$ if (a_n) is the itinerary of x_0, i.e., $T^n x_0$ is represented as a functional of an absolutely regular stochastic process with exponentially decaying mixing coefficients. Furthermore, in this case Φ is Hölder-continuous; i.e., there is some $q<1$ such that $|\Phi(a_0, a_1, a_2,...) - \Phi(b_0, b_1, b_2,...)| \leqslant \text{const.}\, q^k$ if $a_0 = b_0,..., a_k = b_k$.

For the Kaplan–Yorke model T (1.3), how to reduce the problem of representing $(T^n(x_0, y_0))$ as a functional of an absolutely regular process with exponentially decaying mixing coefficients to the corresponding problem for the one-dimensional transformation $x \to 2x \pmod 1$ described above has been shown.[27] Thus, this example fits into the general situation.

At this point we would like to point out that the interval transformations $\hat{T}(x) = 1 - |2x - 1|$ and $R(x) = 1 - |\beta(x - \alpha)|$, $(\beta = (1 + \sqrt{5})/2$, $\alpha = 1 - \beta^{-1})$ also fit into our general framework. The same is true for certain diffeomorphic images of $\hat{T}$ and their perturbations as described in Section 4.

3. *U*-STATISTIC ESTIMATORS FOR DYNAMICAL SYSTEMS

The computations described in Section 4 need some theoretical justification. The theorems available in the literature do not suffice, and we therefore present what is needed in this section. The proofs will be given in Appendix II. We also would like to mention that the theorem presented below has almost immediate extensions to more complicated situations. For example, given two sets of observations arising possibly from two different systems, a similar method can be developed to detect this difference.

We begin with the definition of (one sample) U statistics. Let $X_1,..., X_n$ be random variables with values in $\mathbb{R}^d$, let m be an integer $\geqslant 1$, and let

$$h\colon \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{m \text{ times}} \to \mathbb{R}$$

be a symmetric (measurable) function. Then the associated U statistic is defined by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n} h(X_{i_1,...,} X_{i_m}) \tag{3.1}$$

In this note the $X_1, X_2,...$ will always form a stationary sequence. The simplest example of a U statistic is given by the sample mean $1/n \sum_{i=1}^n X_i$ with $m = 1$, $d = 1$, and $h(x) = x$. Other examples are the sample variance ($m = 3$, $d = 1$, and h being the symmetrization of $(x, y, z) \mapsto (x - y)(x - z)$), the sample covariance for pairs of observations ($m = 3$, $d = 2$, and h being the symmetrization of $((x_1, x_2), (y_1, y_2), (z_1, z_2)) \mapsto (x_1 - y_1)(x_2 - z_1)$), the Wilcoxon statistic ($m = 2$, $d = 1$, $h(x, y) = 1_{\{x + y > 1\}}$), and the statistic $C_n(r)$ in (1.2).

Next we describe the stationary process $X_1, X_2, \dots$ to which the U statistics will be applied. Let $(Z_n : n \geq 1)$ be an absolutely regular stationary sequence with mixing coefficients $\beta(n)$ satisfying

$$\beta(n)^{\delta/(2+\delta)} = O(n^{-2-\varepsilon}) \qquad \text{for some } \varepsilon, \delta > 0 \tag{3.2}$$

and taking values in a finite set. Note that this will be satisfied in our examples by the itinerary process $(a_n(x))$; in fact, $\beta(n)$ will decrease by an exponential rate. Then $X_1, X_2, \dots$ will be taken to be a Lipschitz functional of $(Z_n : n \geq 1)$, i.e., we assume that there exists a function $f(u_1, u_2, \dots)$ satisfying

$$X_j = f(Z_j, Z_{j+1}, \dots) \qquad \text{for} \qquad j \geq 1 \tag{3.3}$$

and f is Lipschitz-continuous in the sense that there is some $\alpha < 1$ such that

$$|f(z_1, z_2, \dots) - f(z_1', z_2', \dots)| \leq \text{const.} \cdot \alpha^n \tag{3.4}$$

if $z_1 = z_1', \dots, z_n = z_n'$.

Again it is not hard to verify this condition in all our examples, because $a_n(x)$ takes values in a finite set and x can be written as a Lipschitz-continuous function of its itinerary provided T is hyperbolic with respect to the Euclidean distance.

We also need to impose some regularity conditions on the kernel functions h. Two types of restrictions are described in the following, and we note that all examples from Section 4 belong to one of these classes.

Class A. There are $L > 0$, $r \geq 0$, and $\rho > 0$ such that

$$|h(x_1, \dots, x_m) - h(y_1, \dots, y_m)| \leq L \cdot \sum_{i,j=1}^{m} \|x_i - y_i\|^\rho (1 + \|x_j\|^r + \|y_j\|^r) \qquad (x_i, y_j \in \mathbb{R}^d) \tag{3.5}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. In this case we say that h satisfies the Lipschitz condition.

Class B. For $h: (\mathbb{R}^d)^m \to \mathbb{R}$, $x \in \mathbb{R}$, and $\varepsilon > 0$, define

$$\operatorname{osc}(h, \varepsilon, x) = \sup\{|h(y) - h(y')| : \|y - x\|, \|y' - x\| < \varepsilon, \; y, y' \in (\mathbb{R}^d)^m\}$$

As a function of x, $\operatorname{osc}(h, \varepsilon, \cdot)$ is measurable, even semicontinuous, and with respect to a probability P on $\mathbb{R}^d$ let us define[20,31]

$$\operatorname{osc}(h, \varepsilon) = \int \operatorname{osc}(h, \varepsilon, x)\, dP^m(x)$$

This function is called the mean oscillation of h. If $\lim_{\varepsilon \to 0} \operatorname{osc}(h, \varepsilon) = 0$, then h is P^m Riemann-integrable, and if $\sup_{\varepsilon > 0} \varepsilon^{-1} \operatorname{osc}(h, \varepsilon) < \infty$, then h is of bounded P^m variation. Here we make an assumption stronger than Riemann integrability but weaker than bounded variation

$$M = \sup_{\varepsilon > 0} \varepsilon^{-r} \operatorname{osc}(h, \varepsilon) < \infty \qquad \text{for some} \qquad r > 0 \tag{3.6}$$

In this case we say that h satisfies the variation condition. At this point we remark that the following lemma is easily verified:

Lemma 1. Let h assume only finitely many values and let the set of discontinuity points of h be contained in a finite union of hyperplanes of $(\mathbb{R}^d)^m$. If P is absolutely continuous with marginals P_j satisfying

$$P_j([a, b]) \leqslant C|b-a|^r \quad \text{for some } r > 0 \text{ and } C > 0 \quad (a, b \in \mathbb{R}) \tag{3.7}$$

then (3.6) holds.

We need one more notation in order to formulate the main result. Let h be as before, and denote by dF the marginal distribution of the stationary process $(X_n : n \geqslant 1)$, i.e., the probability that X_1 belongs to the set $A \subseteq \mathbb{R}^d$ is just $F(A)$. Define a new kernel function h_1 by

$$h_1(x) = \int \cdots \int h(x, x_2, \ldots, x_m)\, dF(x_2) \cdots dF(x_m) \tag{3.8}$$

Denoting by E_F expectation with respect to the distribution F, let

$$\theta = E_F[h_1(X_1)] = \int h_1(x)\, dF(x) \tag{3.9}$$

and

$$\sigma^2 = m^2 \cdot \left(E_F[h_1(X_1)^2] - \theta^2 + \sum_{t \geqslant 2} E_F[(h_1(X_1) - \theta)(h_1(X_t) - \theta)] \right) \tag{3.10}$$

Theorem 1. The sequence U_n of U-statistics converges to θ in probability, and the sequence $\sqrt{n}\,(U_n - \theta)$ converges weakly to the normal distribution $\mathcal{N}(0, \sigma^2)$ with expectation 0 and variance σ^2, provided one of the following conditions (a) or (b) holds:

(a) (i) h satisfies the Lipschitz condition (3.5)

(ii) $(Z_n : n \geqslant 1)$ satisfies the mixing condition (3.2)

(iii) $(X_n : n \geqslant 1)$ is a Lipschitz functional of $(Z_n : n \geqslant 1)$ in the sense of (3.3) and (3.4)

(b) (i) h satisfies the variation condition (3.6)

(ii) $(Z_n : n \geqslant 1)$ satisfies the mixing condition (3.2)

(iii) $(X_n : n \geqslant 1)$ is a Lipschitz functional of $(Z_n : n \geqslant 1)$ in the sense of (3.3) and (3.4)

In all examples considered in this paper it is not hard to see that these conditions hold (use Lemma 1). A proof of this theorem can be found in Appendix II.

Let us now see how Theorem 1 applies. Suppose we want to test for symmetry of distributions. Then it is well-known that the mean or the Wilcoxon statistic perform well in case the $(X_n : n \geqslant 1)$ are independent. If we want to apply these statistics in the present situation, we have to get a good estimate of σ^2 which, contrary to the independent case, involves variance and covariance estimates. The latter are derived again using U statistics with the appropriate kernel functions applied to $(\hat{h}_1(X_k) : k = 1, \ldots, n)$, where

$$\hat{h}_1(X_k) = \binom{n-1}{m-1}^{-1} \sum_{\substack{1 \leqslant i_1 < \cdots < i_{m-1} \leqslant n \\ i_\nu \neq k}} h(X_k, X_{i_1}, \ldots, X_{i_{m-1}})$$

is again a U statistic.

If two or more test procedures for the same problem are available, their relative asymptotic efficiency tells us which of them is preferable in the sense that it needs less observations to produce a decision on a fixed level of reliance than the others. For example, let $f(\theta, x)$ be a family of probability densities on $\mathbb{R}$ indexed by some parameter θ, $|\theta| < \theta_1$, and suppose that f has an asymptotic expansion $f(\theta, x) = f_0(x) + \theta f_1(x) + O(\theta^2)$. Each $f(\theta, x)$ might be the density of an invariant measure of a dynamical system that we are observing, and we must decide on the basis of n successive observations $X_1 = x$, $X_2 = Tx, \ldots, X_n = T^{n-1}x$ whether $\theta = 0$ (H_0) or not (H_1).

Suppose $S_n = S_n(X_1, \ldots, X_n)$ are statistics which converge in probability to a number $s(\theta)$ if the underlying invariant density is $f(\theta, x)$. Intuitively

$$\operatorname{eff}(S_n) = \left(\frac{ds}{d\theta}\right)^2_{\theta=0} \Big/ \operatorname{var}(S_n)$$

is a good measure of performance of S_n. If S_n and S'_n are two different statistics, both designed for testing H_0 against H_1 and both asymptotically normal (when suitably normalized), then

$$Q = \frac{\mathrm{eff}(S_n)}{\mathrm{eff}(S'_n)}$$

is their relative asymptotic efficiency, and the classical theorem of Pitman and Noether says that S_n performs better than S'_n (in the sense described above) if $Q > 1$. In other words, given a number n, choose a (minimal) number $m = m_n$ such that the procedures based on S_n and S'_m perform equally well. Then $\lim_{n \to \infty} m_n/n = Q$. Thus we see how Theorem 1 applies to efficiency comparisons.

4. EXAMPLES

I. Tests for Symmetry of the Invariant Densities of Interval Maps

One of the simplest dynamical system is the so-called "hut-map" $\hat{T}: [0, 1] \to [0, 1]$, $\hat{T}(x) = 1 - |2x - 1|$ (mapping T_1 from the introduction). The constant function 1 is its unique invariant density. If $u: [0, 1] \to [0, 1]$ is any diffeomorphism with $u' \geqslant 0$, one can construct from $\hat{T}$ the map $T_u(x) = u^{-1}(\hat{T}[u(x)])$, and it is not hard to see that $u'(x)$ is its unique invariant density. Hence, if $u'(x)$ is symmetric at $\frac{1}{2}$ (equivalently if $u(1-x) = 1 - u(x)$), T_u is an interval map with symmetric invariant density.[13] Denote this class of conjugating diffeomorphisms by U. We study perturbations $T_u(\varepsilon, x) = T_u(x) + \varepsilon G(T_u x)$, where G is symmetric at $\frac{1}{2}$, analytic, and $0 \leqslant y + G(y) \leqslant 1$ $(0 \leqslant y \leqslant 1)$. For small ε, the invariant density $h_u(\varepsilon, \cdot)$ of $T_u(\varepsilon, \cdot)$ is analytic in ε, i.e.

$$\begin{aligned} &h_u(\varepsilon, x) = u'(x) + \varepsilon h_u^1(x) + \varepsilon^2 h_u^2(x) + \cdots \qquad \text{where} \\ &h_u^1 = -(u'G)' \qquad \text{is antisymmetric} \end{aligned} \tag{4.1}$$

(Here we have to assume that u^{-1} is analytic with nonvanishing derivative except at $z = 0$ and $z = 1$, where $(u^{-1})'(z) = 0$, $(u^{-1})''(z) \neq 0$ are allowed (see Ref. 19).) Hence $h_u(\varepsilon, \cdot)$ is symmetric if and only if $\varepsilon = 0$, and tests for symmetry can be based on the following functionals:
Expectation

$$\mu(u, \varepsilon) = \int x h_u(\varepsilon, x)\, dx = \tfrac{1}{2} + \varepsilon \int u'(x)\, G(x)\, dx + 0(\varepsilon^2) \tag{4.2}$$

Wilcoxon functional

$$w(u, \varepsilon) = \iint 1_{\{x+y<1\}} h_u(\varepsilon, x)\, h_u(\varepsilon, y)\, dx\, dy$$

$$= \tfrac{1}{2} + 2\varepsilon \int [u'(y)]^2\, G(y)\, dy + 0(\varepsilon^2)$$

(use (4.1), the symmetry properties and integration by parts).

The corresponding U statistics

$$\hat{S}_N(x) = \frac{1}{N} \sum_{i=0}^{N-1} T^i(x) \tag{4.4}$$

$$\hat{W}_N(x) = \frac{1}{N(N-1)} \sum_{\substack{i,j=0\\ i\neq j}}^{N-1} 1_{\{T^ix + T^jx > 1\}} \tag{4.5}$$

where T is one of the transformations $T_u(\varepsilon, \cdot)$ ($u \in U$, ε small) are asymptotically unbiased estimators of $\mu(u, \varepsilon)$ and $w(u, \varepsilon)$, respectively. A direct computation using the symmetry of T_u and u' shows that $T_u^k x$ and $T_u^l x$ are uncorrelated if $k \neq l$. Hence

$$\lim_{N\to\infty} \operatorname{var}_{u,\varepsilon}[\sqrt{N}\, \hat{S}_N] = \int (x - \tfrac{1}{2})^2\, u'(x)\, dx + 0(\varepsilon)$$

while

$$\lim_{N\to\infty} \sqrt{N}\, \{E_{u,\varepsilon}[\hat{S}_N] - \mu(u, \varepsilon)\} = 0$$

and

$$\frac{\partial}{\partial\varepsilon} \mu(u, \varepsilon)\bigg|_{\varepsilon=0} = \int u'(x)\, G(x)\, dx$$

such that by Noether's theorem the asymptotic efficacy of $\hat{S}_N$ is

$$\operatorname{eff}(\hat{S}_N) = \left[\int u'(x)\, G(x)\, dx\right]^2 \Big/ \int (x - \tfrac{1}{2})^2\, u'(x)\, dx \tag{4.7}$$

using the notation of Chapter 5 in Ref. 35. Similarly one shows that

$$\operatorname{eff}(\hat{W}_N) = 12\left[\int u'(x)^2\, G(x)\, dx\right]^2 \tag{4.8}$$

For $G(x) = x(1-x)$ (which produces $\hat{T}(\varepsilon, x) = (1-\varepsilon)\, \hat{T}(x) + \varepsilon 4x(1-x)$) and for various choices of u' the numerical values are displayed in Table I.

Table I.

u'	1	$\frac{1}{\pi\sqrt{x(1-x)}}$	$\alpha+\beta(x-\frac{1}{2})^{2n}$ $(0\leqslant\alpha\leqslant 1,\quad n\geqslant 1,\quad \beta=(1-\alpha)(2n+1)\,2^{2n})$
$\mathrm{eff}(\hat{S}_N)$	$\frac{1}{3}$	$\frac{1}{8}$	$\left(\frac{\alpha}{6}+\frac{1-\alpha}{4n+6}\right)^2\Big/\left(\frac{1}{4}-\frac{\alpha}{6}-\frac{1-\alpha}{4n+6}\right)$
$\mathrm{eff}(\hat{W}_N)$	$\frac{1}{3}$	$\frac{12}{\pi^4}$	$12\left(\frac{\alpha^2}{6}+\frac{\alpha(1-\alpha)}{2n+3}+(1-\alpha)^2\frac{(2n+1)^2}{2(4n+1)(4n+3)}\right)^2$
$I=\int\frac{(h_u^1)^2}{u'}$	$\frac{1}{3}$	$\frac{1}{8}$	?
$Q=\frac{\mathrm{eff}(\hat{S}_N)}{\mathrm{eff}(\hat{W}_N)}$	1	~ 1.015	

Remarks. (a) $u'(x)=1/\pi\sqrt{x(1-x)}$ is the invariant density of $T_u(x)=4x(1-x)$. (b) It is straightforward to show that $\mathrm{eff}(\hat{U}_N)\leqslant I=\int[(\partial/\partial\theta)\,p(\theta,x)]^2/p(\theta,x)\,dx$ if $\hat{U}_N$ is a U statistic based on samples x, $T_\theta x$, $T_\theta^2 x,\ldots,T_\theta^{N-1}x$ for a family $\{T_\theta\colon|\theta|<\theta_0\}$ of transformations with invariant densities $p(\theta,x)$. (c) For $u'(x)=\alpha+\beta(x-\frac{1}{2})^{2n}$ Fig. 3 shows that for small

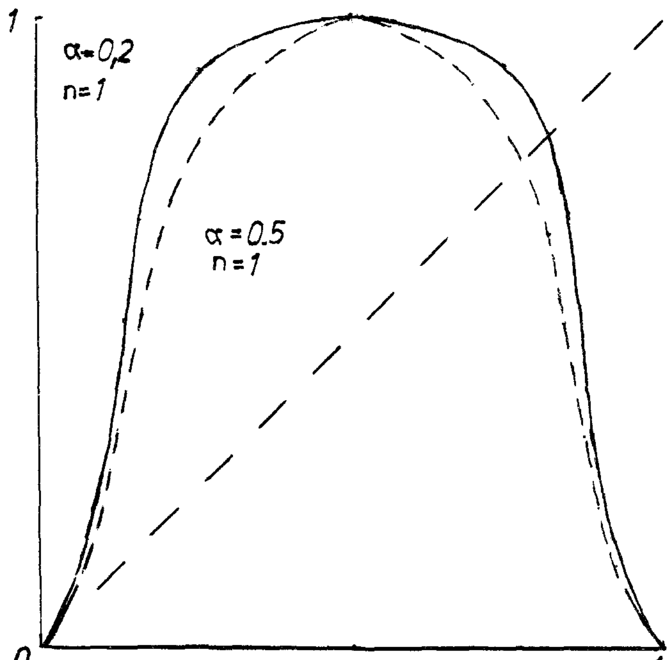

Figure 2

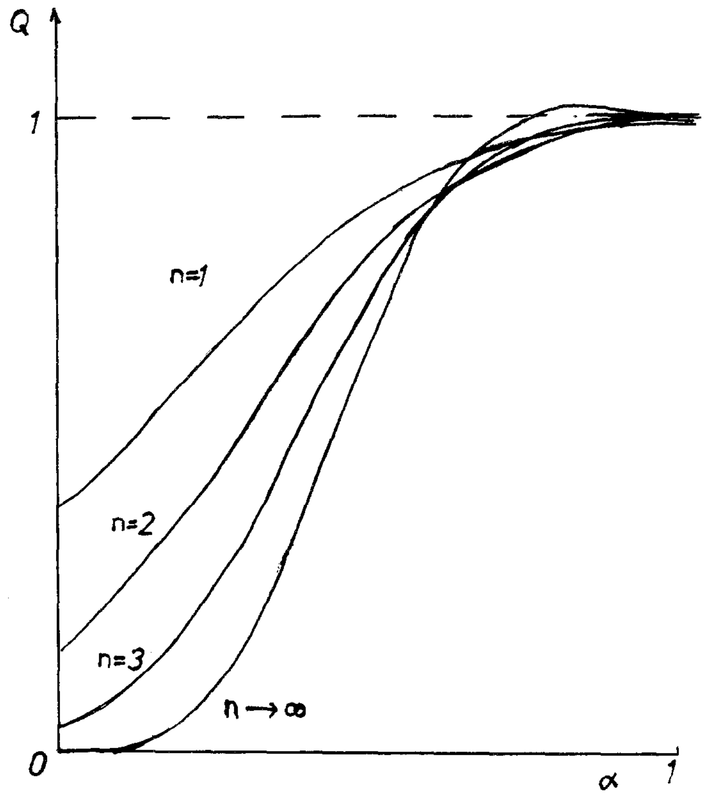

Figure 3

Table II.

ε	$\hat{S}_N$	$\mu(1, \varepsilon)$	$\hat{\sigma}$	σ	$\hat{W}_N$	$w(1, \varepsilon)$	$\hat{\sigma}$	σ
−0.2	0.4744	0.4590	0.0075	0.0077	0.4478	0.4138	0.0153	0.0152
−0.2	0.4583	0.4590	0.0077	0.0077	0.4180	0.4138	0.0152	0.0152
−0.1	0.4660	0.4815	0.0079	0.0077	0.4343	0.4632	0.0153	0.0154
−0.1	0.4945	0.4815	0.0075	0.0077	0.4887	0.4632	0.0154	0.0154
0	0.5053	0.5	0.0076	0.0077	0.5108	0.5	0.0154	0.0154
0	0.4959	0.5	0.0078	0.0077	0.4921	0.5	0.0154	0.0154
0.1	0.5223	0.5151	0.0076	0.0077	0.5452	0.5302	0.0154	0.0154
0.1	0.5212	0.5151	0.0076	0.0076	0.5016	0.5302	0.0153	0.0154
0.2	0.5338	0.5276	0.0076	0.0076	0.5678	0.5546	0.0153	0.0153
0.2	0.5380	0.5276	0.0076	0.0076	0.5769	0.5546	0.0152	0.0153

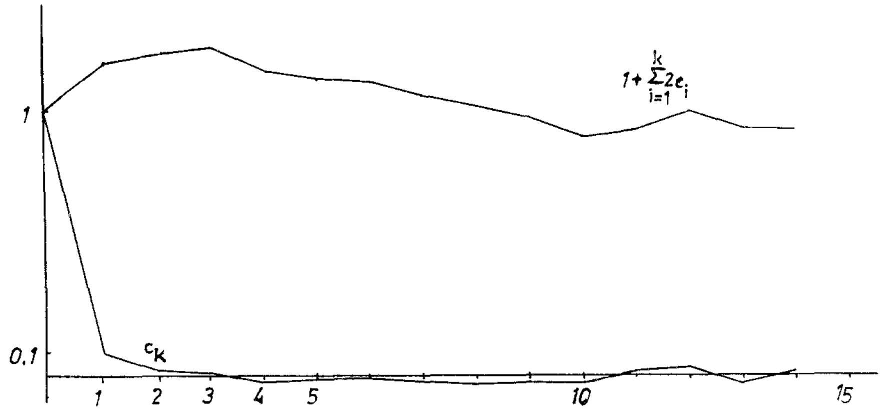

Figure 4

n, $\hat{W}_N$ is in general asymptotically more efficient than $\hat{S}_N$, while for large n and α close to 0.8, $\hat{S}_N$ is slightly better than $\hat{W}_N$. (Fig. 2 shows the shapes of some of these transformations T_u.)

The results of numerical simulations were in good agreement with the asymptotic theory for $N \geqslant 250$ (roughly). In Table II we summarize some results when $u'(x) = 1$, $N = 1400$, $\varepsilon = j/10$ ($j = -2, \ldots, 2$). For each ε we used two different initial values.

If we had not known that the observations are uncorrelated, this could have been easily deduced from a correlogram, which for $\hat{S}_N$ typically looks that shown in Fig. 4. A completely different correlogram can be found in the next example.

II. An Example with Important Correlations

Due to the particular symmetries of the transformations T_u, the asymptotic variances of the test statistics in (I) were the same as if the underlying observations had been independent. Here we give a simple example where the correlations reduce the asymptotic variance to nearly 7.5% of that in the independent case.

Let $R: [0, 1] \to [0, 1]$ be given by $R(x) = 1 - |\beta(x - \alpha)|$, where $\beta = (1 + \sqrt{5})/2$ is the positive solution of $\beta^2 - \beta - 1 = 0$ and $\alpha = 1 - \beta^{-1}$. Then $R: 0 \to \alpha \to 1 \to 0$ (see Fig. 5), and the simple orbit structure of the critical points of R permits the explicit computation of the following quantities:

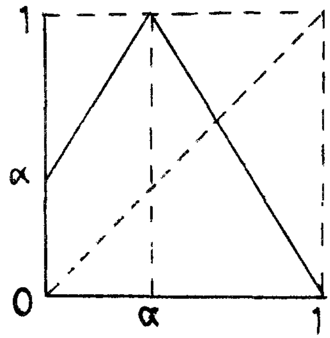

Figure 5

The unique invariant density $h(x)$ of R is

$$h(x)=\begin{cases}(1+\alpha)^{-1} & \text{if } x\leqslant\alpha\\ \beta(1+\alpha)^{-1} & \text{if } x>\alpha\end{cases}$$

$$\int xh(x)\,dx=\tfrac{2}{5}(1+\alpha), \qquad \left(\int xh(x)\,dx\right)^2=\frac{4}{5\beta^2} \tag{4.9}$$

$$\int x^2h(x)\,dx=\frac{1}{\beta^2}$$
$$c_0:=\int x^2h(x)\,dx-\left(\int xh(x)\,dx\right)^2=\frac{1}{5\beta^2}\sim 0.0764 \tag{4.10}$$

$$c_k:=\int xR^k(x)\,h(x)\,dx-\left(\int xh(x)\,dx\right)^2$$
$$=\frac{c_0}{\beta^{2n}}\left((-1)^n-\frac{\sqrt{5}}{3}\,\varepsilon_n\right) \tag{4.11}$$

where $\varepsilon_n=0$ if $n=0$ mod 3, 1 if $n=1$ mod 3, and -1 if $n=2$ mod 3.

Hence the asymptotic variance of the mean-value estimator $\hat{S}_N(x)$ is

$$\sigma^2=c_0+2\sum_{k=1}^{\infty}c_k=\frac{c_0}{6\sqrt{5}}\sim c_0\cdot 0.0745\sim 0.0057 \tag{4.12}$$

The calculations are based on the knowledge of some of the eigenfunctions of the Perron–Frobenius operator $\mathscr{L}$ of R, $\mathscr{L}f(x)=\beta^{-1}\sum_{Ry=x}f(y)$

$$\mathscr{L}h=h$$

$$\mathscr{L}e_2=(\beta-2)\,e_2 \qquad \text{with } e_2(x)=\begin{cases}1 & \text{if } x\leqslant\alpha\\ 1-\beta & \text{if } x>\alpha\end{cases}$$

$$\mathscr{L}e_{3,4} = (2-\beta)\,\gamma_\pm e_{3,4} \qquad \text{with } \gamma_\pm = (-1 \pm i\sqrt{3})/2 \qquad \text{and}$$

$$e_{3,4} = \begin{cases} 1 + (\gamma_\pm \beta - 1)\,x & \text{if } x \leqslant \alpha \\ -(2\gamma_\pm + \beta) + [\gamma_\pm(\beta+1)+\beta]\,x & \text{if } x > \alpha \end{cases}$$

(For further details on this operator, see Appendix I.)

Here are three values of $\hat{S}_N(x)$ for $N = 1400$. The second value in each row gives the standard deviation estimated from the data, and the third one the number of correlations that have been taken into account

0.5508	0.0028	5
0.5516	0.0039	5
0.5534	0.0020	5

These results should be compared with the theoretical values

$$0.5528 \qquad 0.0020$$

Figure 6 shows the correlograms of two of the three samples (the third is similar) and Fig. 7 the corresponding partial correlation sums on which the decision was based how many correlations should be considered for the estimation of the asymptotic variance. The difference from Fig. 4 is striking.

For $\beta = 1.5$, a value for which we could not compute the theoretical values as above, the estimation of the asymptotic variance of S_N is still

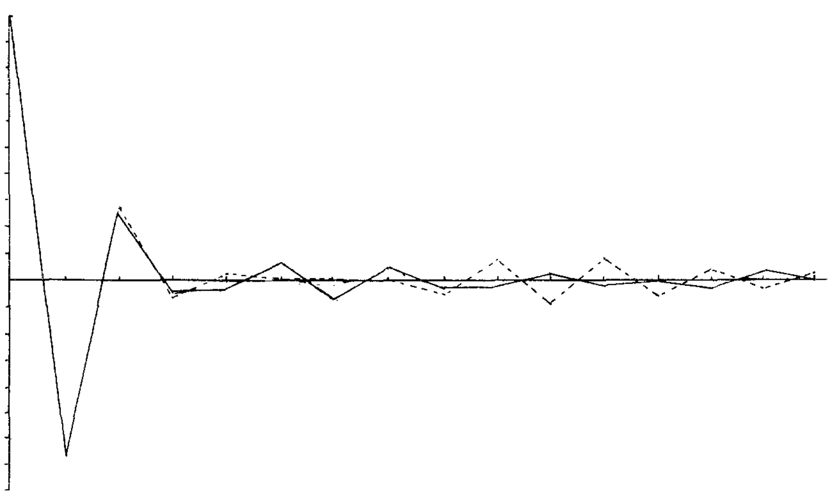

Figure 6

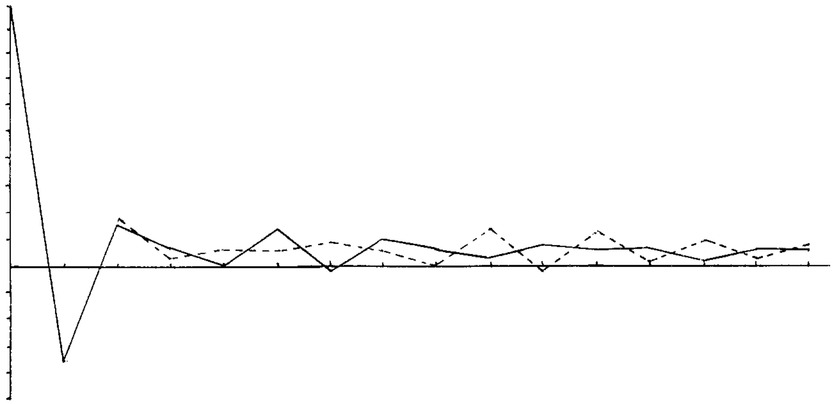

Figure 7

more delicate, as can be seen from Fig. 8 where the partial correlation sums are plotted against time ($N = 1400$). Here are three values of $S_N(x)$ with estimated standard deviations and number of covariances taken into account

0.5509	0.0018	14
0.5531	0.0031	10
0.5529	0.0018	14

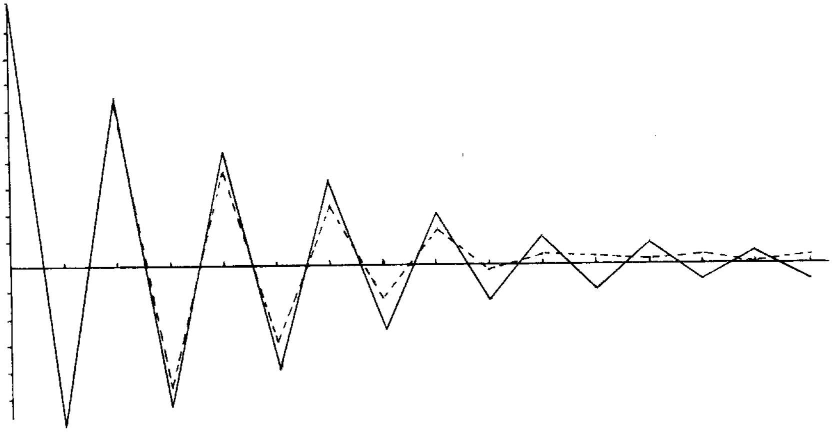

Figure 8

III. Estimation of the Grasberger–Procaccia Correlation Exponent

We consider the dynamical system (1.3) with $\lambda = 0.2$ and estimate the vector $[C(r_i)]$ with five values of $r_i = 0.08 \cdot 2^{-i}$ using the estimator $[\hat{C}_N(r_i)]$—cf. (1.1) and (1.2). Here and in the sequel, the index i ranges from 0 to 4. The distribution of $\{\sqrt{N}\,[\hat{C}_N(r_i) - C(r_i)]\}$ approaches a five-dimensional normal distribution ($N \to \infty$) with zero expectation and a certain covariance matrix (σ_{ij}^2), which can be consistently estimated from the underlying data. The normality result is a variant of Theorem 1. In exceptional cases the distribution might be degenerate. We consider this problem later.

Assuming that $C(r)$ obeys a power law, i.e., $C(r) \simeq Kr^{\nu}$ $(r \to 0)$ for some constants K and ν, the vectors $[\log \hat{C}_N(r_i)]$ $(N \in \mathbb{N})$ provide a consistent and asymptotically unbiased sequence of estimators for

$$[\log C(r_i)] \simeq [\log K + \nu \log r_i] \tag{4.13}$$

Furthermore, $\sqrt{N}\,[\log \hat{C}_N(r_i) - \log C(r_i)]$ tends to a five-dimensional normal distribution with zero expectation and covariance matrix $(c_{ij}) = [C(r_i)^{-1}\,\sigma_{ij}^2 C(r_j)^{-1}]$. Assuming strict equality in (4.13), one obtains a consistent and asymptotically unbiased sequence of least-squares estimators $\hat{\nu}_N$ for ν, where $\sqrt{N}\,(\hat{\nu}_N - \nu)$ is asymptotically normal with variance σ^2. More exactly

$$\hat{\nu}_N = \frac{1}{5} \sum_i u_i \log \hat{C}_N(r_i), \qquad \sigma^2 = \frac{1}{5} \sum_i \frac{1}{5} \sum_j u_i c_{ij} u_j$$

where

$$u_i = \left(\log r_i - \frac{1}{5} \sum_k \log r_k \right) \Bigg/ \left[\frac{1}{5} \sum_k (\log r_k)^2 - \left(\frac{1}{5} \sum_k \log r_k \right)^2 \right]$$

σ^2 can be consistently estimated since the c_{ij} can.

In Table III we give the results of 20 runs with $N = 750$. $\hat{\nu}_N$ is the estimated value for ν, $\hat{\sigma}_4$ is the estimated standard deviation with four covariances taken into account, $\hat{\sigma}_0$ (the estimated standard deviation without covariances) is given for comparison, and $\hat{\Delta}$ is the mean square deviation of $[\log \hat{C}_N(r_i)]$ from the straight line $\log C(r) = \log \hat{K}_N + \hat{\nu}_N \log r$.

These results should be compared to the value $D = 1 + (\log 2/\log 5) \simeq 1.431$ for the Hausdorff dimension of the system's attractor[7] which is close

Table III.

$\hat{v}_N$	$2\hat{\sigma}_4$	$2\hat{\sigma}_0$	$2\hat{\Delta}$
1.378	0.054	0.037	0.019
1.398	0.049	0.035	0.045
1.442	0.053	0.036	0.017
1.374	0.043	0.033	0.033
1.438	0.045	0.034	0.027
1.419	0.033	0.032	0.017
1.409	0.041	0.032	0.012
1.436	0.041	0.033	0.064
1.379	0.059	0.040	0.038
1.444	0.040	0.034	0.027
1.465	0.047	0.036	0.031
1.397	0.034	0.032	0.042
1.449	0.042	0.036	0.029
1.476	0.049	0.035	0.027
1.395	0.058	0.036	0.055
1.376	0.038	0.034	0.026
1.449	0.042	0.034	0.037
1.351	0.053	0.034	0.041
1.444	0.053	0.036	0.024
1.426	0.046	0.033	0.069

to $v(v \leqslant D)$. Grasberger and Procaccia[8] obtained the estimate 1.42 ± 0.01 for v on the basis of $N = 15000$ data, where the error bound is derived from the mean square deviation. Unnecessary to say that with our moderate sample size we did not attempt to obtain a better estimate than Grassberger–Procaccia, but we want to point out how such estimation problems can be treated in a statistically rigorous way. Nevertheless the above data need some comment

1. The estimates for Δ fluctuate much more than those for the σ_k and seem to be a very unreliable estimate for the error bound of $\hat{v}_N$.
2. Although the estimate of σ_4 fluctuates much more than that of σ_0, it is evident that a conservative estimation of the error bound for $\hat{v}_N$ cannot neglect the time correlation of the data.

As mentioned earlier, the general theory does not exclude the case of a degenerate limiting distribution (although this is very unlikely to happen). Therefore we checked the approximate normality of the statistic $\sqrt{N}\,(\hat{v}_N - v)/\hat{\sigma}_4$: Let v_{med} denote the median of the above tabulated 20 values of $\hat{v}_N$. For the statistic $\sqrt{N}\,(\hat{v}_N - v_{\text{med}})/\hat{\sigma}_4$ we calculated the value of

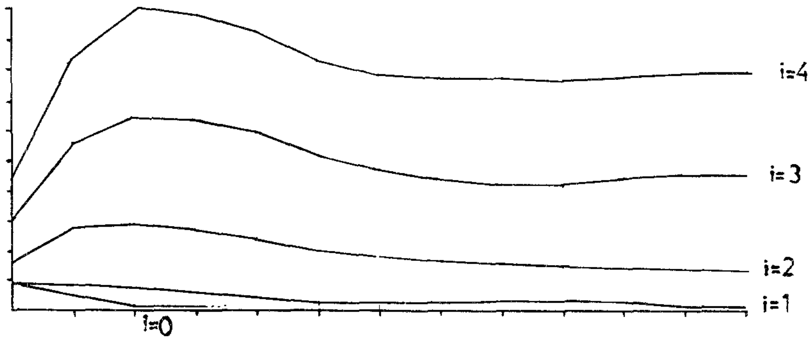

Figure 9

the Shapiro–Wilk test on normality as $W = 0.95$. (The 10% and 90% levels are 0.920 and 0.979, respectively.) Five further tests on groups of 50 independently calculated $\hat{v}_N$ values also supported the assumption of normality.

As a final remark, consider Fig. 9. It shows the partial correlation sums for the estimators $\log \hat{C}_N(r_i)$ ($i = 0, \ldots, 4$; $N = 750$). It is clearly seen that even very similar estimators (based on the same set of data) may produce rather different correlation coefficients, in particular correlations between the data may both augment and reduce the asymptotic variance of estimators.

APPENDIX I

The Perron–Frobenius Operator

Let $(X, \mathscr{B}, m)$ be a nonatomic probability space, $T: X \to X$ a measurable, nonsingular, and noninvertible transformation. The Perron–Frobenius operator $\mathscr{L}: L_m^1 \to L_m^1$ is defined by

$$\int f \cdot (g \circ T)\, dm = \int \mathscr{L}(f) \cdot g\, dm \qquad \text{for all } f \in L_m^1 \text{ and } g \in L_m^\infty$$

In words, $\mathscr{L}(f)$ can be described as follows: If the system at time t is described by the distribution $f\, dm$, then it is described at time $t+1$ by $\mathscr{L}(f)\, dm$. If T' denotes the Radon–Nikodym derivative of T with respect to m, one can write

$$\mathscr{L}(f)(x) = \sum_{Ty = x} f(y)/T'(y)$$

In many cases $\mathcal{L}$ can be restricted to a linear subspace B of L_m^1 which can be endowed with a Banach-space structure in such a way that $\mathcal{L}_{|B}$ is a quasicompact operator; see, e.g., Refs. 15, 16, and 26. For piecewise expanding interval transformations the space of functions of bounded variation is a good choice for B. If the system (T, m) is mixing, i.e., if $m(T^{-n}A \cap B)$ converges as $n \to \infty$ for all $A, B \in \mathcal{B}$, then $\mathcal{L}$ can be decomposed

$$\mathcal{L} = \Phi + \Psi$$

where Φ is a projection on the one-dimensional space spanned by the unique invariant density $h = \Phi(1)$ and where Ψ is a linear operator on B with spectral radius strictly less than 1, $\Phi\Psi = \Psi\Phi = 0$. Using this decomposition one can show that the coefficients of absolute regularity β_k defined in Section 2 decrease exponentially fast to 0. This situation is met in all examples of Section 5.

If (T, m) is not mixing, the quasicompactness of $\mathcal{L}$ still guarantees that it can be decomposed into finitely many ergodic components on each of which (T, m) can be described as a product of a finite cyclic permutation with a mixing system.

APPENDIX II. PROOF OF THEOREM 1

In this appendix we give a proof of Theorem 1 in a slightly more general context, and afterward show that these general conditions are met in our specific situation. We assume throughout that $X_1, X_2,...$ is a (strictly) stationary E-valued process where (E, B) denotes a measurable space.

Let $h: E^m \to \mathbb{R}$ $(m \geqslant 1)$ be a measurable function and let $\Delta \subseteq \mathbb{N}^m$ be defined by $\Delta = \{(t_1,..., t_m): t_i \neq t_j\ (i \neq j)\}$. Let $p \geqslant 1$. We say that the process $(X_n: n \geqslant 1)$ satisfies the condition (A_p) with coefficients $\psi(n) \downarrow 0$ and $\beta(n) \downarrow 0$ (as $n \to \infty$) if there exist an $\eta > 0$, an absolutely regular, strictly stationary process $Z_1, Z_2,...$ with coefficients of absolute regularity $\beta(n)$ and measurable functions f_l $(l \geqslant 0)$ such that

$$X_j = X_j^0 = f_0(Z_j, Z_{j+1},...) \qquad (j \geqslant 1) \tag{II.1}$$

Setting $X_j^l = f_l(Z_j, Z_{j+1},..., Z_{j+l-1})$ we have for every $(t_1,..., t_m) \in \Delta$

$$\begin{aligned}(E|h(X_{t_1},..., X_{t_m}) - h(X_{t_1}^l,..., X_{t_m}^l)|^p)^{1/p} \\ \leqslant \psi(\max\{1, \min_{1 \leqslant i \neq j \leqslant m} |t_i - t_j| - l\}) + l^{-1-\eta}\end{aligned} \tag{II.2}$$

Moreover, we say that $(X_n: n \geqslant 1)$ satisfies the condition (σ_δ) where $\delta \geqslant 0$ if

$$\sigma_\delta^{2+\delta} = \sup_{(t_1,...,t_m) \in \Delta} E|h(X_{t_1},..., X_{t_m})|^{2+\delta} < \infty$$

We first note the following:

Lemma 2. Let $(X_n : n \geqslant 1)$ satisfy conditions $(A_{2+\delta})$ and (σ_δ). Let $(X_n^l(v) : n \geqslant 1)$ $(v = 1,..., m)$ denote independent copies of the process $(X_n^l : n \geqslant 1)$ $(l \geqslant 1)$. Then

$$\sup[E|h(X_{t_1}^l(v_1),..., X_{t_m}^l(v_m)|^{2+\delta}]^{1/(2+\delta)} \leqslant \sigma_\delta + \psi(1) + 1 \tag{II.3}$$

where the supremum extends over $(t_1,..., t_m) \in \Delta$, $l \geqslant 1$, and $(v_1,..., v_m) \in \{1,..., m\}^m$.

Proof. Fix $v \in \{1,..., m\}^m$, $l \geqslant 1$, and $(t_1,..., t_m) \in \Delta$. Using Lemma 8 in Ref. 5 it follows that

$$[E|h(X_{t_1}^l(v_1),..., X_{t_m}^l(v_m))|^{2+\delta}]^{1/(2+\delta)}$$
$$\leqslant \sup_{(s_1,...,s_m) \in \Delta} [E|h(X_{s_1}^l,..., X_{s_m}^l)|^{2+\delta}]^{1/(2+\delta)}$$

Applying (II.2) and taking the supremum on the left-hand side, (II.3) follows.

The function h is called degenerate if

$$Eh(a_1,..., a_{t-1}, X_1, a_t,..., a_{m-1}) = 0 \tag{II.4}$$

for all $1 \leqslant t \leqslant m$ and $a_1,..., a_{m-1} \in E$.

Lemma 3. Let $m \geqslant 2$. Let $(X_n : n \geqslant 1)$ satisfy condition $(A_{2+\delta})$ and (σ_δ) for some $\delta > 0$ such that

$$\beta(n)^{\delta/(2+\delta)} = O(n^{-2-\varepsilon}) \qquad \text{for some} \qquad 0 < 2\varepsilon < \eta \qquad \text{and} \tag{II.5}$$

$$\sum_{k \geqslant 1} \psi(k) < \infty \tag{II.6}$$

If h is degenerate, then

$$ER_N(h)^2 = o(N^{2m-1}) \qquad \text{where} \tag{II.7}$$

$$R_N(h) = \sum_{\substack{(t_1,...,t_m) \in \Delta_N \\ 1 \leqslant t_j \leqslant N}} h(X_{t_1},..., X_{t_m}) \tag{II.8}$$

Proof. Denote by $R_N^l(h)$ the random variable defined as in (II.8) replacing the (X_n) by (X_n^l). By elementary manipulations and setting $l = [N^{1/2-\eta/4}]$ we have that

$$| \|R_N(h)\|_{2+\delta} - \|R_N^l(h)\|_{2+\delta}|$$
$$= O\left\{N^{m-1}\left[l + \sum \psi(k)\right] + N^m l^{-1-\eta}\right\}$$
$$= o(N^{m-(1/2)}) \tag{II.9}$$

$(X_n^l: n \geqslant 1)$ is an absolutely regular sequence with mixing coefficients $\beta((n-l) \vee 1)$. By Lemma 2 all $2+\delta$ moments are uniformly bounded, and hence Proposition 2 of Ref. 5 applies

$$ER_N^l(h)^2 \leqslant \Gamma l^2 N^{2m-2+\varepsilon} \tag{II.10}$$

where Γ denotes a constant depending only on $[\beta(n): n \geqslant 1]$, σ_δ, ε, and δ. (Note that (II.10) is slightly different from the assertion in Proposition 2, but inspecting its proof the statement follows easily.) Since, by the choice of l

$$l^2 N^{-2+\varepsilon} = o(N^{-1})$$

(II.7) follows easily.

In the proof of the last lemma the $(A_{2+\delta})$ condition was needed in order to apply Lemma 2 and Proposition 2, but for the latter some (A_p) condition $(p \geqslant 1)$ would suffice. Consequently, if the conclusion of Lemma 2 holds for other reasons, some (A_p) condition suffices. This yields:

Lemma 4. Let $m \geqslant 2$. Let $(X_n: n \geqslant 1)$ satisfy condition (A_p) for some $p \geqslant 1$ such that (II.5) and (II.6) hold for some $\delta > 0$ and $0 < 2\varepsilon < \eta$. If h is degenerate and bounded, then

$$[E|R_N(h)|^q]^{1/q} = o(N^{m-1/2}) \tag{II.11}$$

where $q = \min(p, 2)$. The convergence in (II.11) is uniform over bounded families of functions h.

Define

$$h_j(x_1,..., x_j) = \int \cdots \int h(x_1, x_2,..., x_m) \prod_{k=j+1}^{m} dP(x_k) \quad (x_1,..., x_j \in E) \tag{II.12}$$

where P denotes the distribution of X_1 and, as in (3.9)

$$\theta = Eh_1(X_1) \tag{II.13}$$

We have

Theorem 2. The sequence $\sqrt{N}\,(U_N - \theta)$ $(N \geqslant 1)$ converges weakly to the normal distribution $N(0, \sigma^2)$ with expectation 0 and variance

$$\sigma^2 = m^2 \left\{ Eh_1(X_1)^2 + 2 \sum_{t \geqslant 2} Eh_1(X_1)\, h_1(X_t) \right\} \tag{II.14}$$

provided one of the following conditions is satisfied:

(a) $(X_n: n \geqslant 1)$ satisfies the conditions (σ_δ) and $(A_{2+\delta})$ w.r.t. each h_j $(j = 1, \ldots, m)$ for some $\delta > 0$ such that (II.5) and (II.6) hold.

(b) $(X_n: n \geqslant 1)$ satisfies the condition (A_p) for some $p \geqslant 1$ such that (II.5) and (II.6) hold with respect to each h_j $(1 \leqslant j \leqslant m)$ and h is bounded.

Proof. According to Hoeffding's decomposition method we may write

$$U_N = \sum_{c=0}^{m} \binom{m}{c} U_N^c$$

where U_N^c denotes a U statistic with respect to a symmetric degenerate and measurable function $\tilde{h}_c: E^c \to \mathbb{R}$. $\tilde{h}_c$ can be written as a linear combination of the functions h_j, and hence $(X_n: n \geqslant 1)$ satisfies the conditions (σ_δ) and $(A_{2+\delta})$ (resp. (A_p)) with respect to each $\tilde{h}_c$. From Lemmas 3 and 4 we conclude that

$$\sqrt{N} \sum_{c=2}^{m} \binom{m}{c} U_N^c \to 0$$

in probability.

If $c = 0$, then $U_N^c = \theta$ and if $c = 1$ then $\tilde{h}_1 = h_1 - \theta$. Consequently, $\sqrt{N}(U_N - \theta)$ and $(m/\sqrt{N}) \sum_{t=1}^{N} [h_1(X_t) - \theta]$ have the same limit distribution. By assumption, we can apply Theorem 18.6.2 of Ref. 17 to the sequence $\{h_1(X_n): n \geqslant 1\}$ to conclude that

$$\frac{m}{\sqrt{N}} \sum_{t=1}^{N} [h_1(X_t) - \theta] \to N(0, \sigma^2)$$

weakly.

In order to obtain Theorem 1 from Theorem 2, we note that the f from (II.2) are easily constructed, because f is Lipschitz-continuous, and the following lemmas hold:

Lemma 5. Let h satisfy the Lipschitz condition (3.5) and suppose that (X_n) satisfies $\|X_1\|_p < \infty$ and

$$\psi_p(l) = (E|X_1 - X_1^l|^p)^{1/p} \to 0 \text{ as } l \to \infty \text{ for some } p > 2\,(r + \rho) \qquad \text{(II.15)}$$

Then (X_n) satisfies the conditions $(A_{2+\delta})$ and $(\sigma_{2+\delta})$ with respect to each h_j for some $\delta > 0$.

Proof. Let $(2 + \delta)(r + \rho) \leqslant p$. Then

$$\begin{aligned} &\|h(X_{t_1}, \ldots, X_{t_m}) - h(X_{t_1}^l, \ldots, X_{t_m}^l)\|_{2+\delta} \\ &\qquad \leqslant \text{const. } [\psi_p(l)]^\rho\, [1 + 2\,\|X_1\|_p + \psi_p(l)]^r \end{aligned}$$

and, noticing that together with h also each h_j satisfies the Lipschitz condition, the lemma follows immediately.

Lemma 6. Let h satisfy the variation condition (3.6). If f is Lipschitz continuous, then (X_n) satisfies condition (A_1) with respect to each h_j.

Proof. Observe first that with h also each h_j satisfies condition (3.6). Since (3.6) implies that h is almost surely bounded (cf. Lemma 1.4 of Ref. 20), we have

$$\begin{aligned} &E|h_j(X_{t_1},..., X_{t_m}) - h_j(X^l_{t_1},..., X^l_{t_m})| \\ &\quad \leqslant E\{\operatorname{osc}[h_j, C2^{-s}, (X^l_{t_1},..., X^l_{t_j})]\} \\ &\quad \leqslant \int \operatorname{osc}(h_j, C2^{-s}, x)\, dP^j(x) + 8\,\|h\|_\infty\,(j-1)\,\beta((\min|t_i - t_k| - l) \vee 1) \\ &\quad \leqslant MC^r 2^{-rs} + 8\,\|h\|_\infty\,(j-1)\,\beta)(\min_{i \neq k}|t_i - t_k| - l) \vee 1) \end{aligned}$$

(The second inequality follows from Lemma 1 of Ref. 41.) Consequently (II.2) is satisfied with $p = 1$ and $\psi(n) \cong \beta(n)$.

REFERENCES

1. B. Bunow and G. H. Weiss, *Math. Biosci.* **47**:221–237 (1979).
2. R. C. Bradley, *Stoch. Proc. Appl.* **14**:67–77 (1983).
3. I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory. Grundl. der math. Wiss.* 245, (Springer-Verlag, Berlin, 1982).
4. M. Denker, *Mathem. Forsch.* **12**:35–47 (1982).
5. M. Denker and G. Keller, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **64**:505–522 (1983).
6. J. D. Farmer, E. Ott, and J. A. Yorke, *Physica* **7D**:153–180 (1983).
7. P. Frederickson, J. L. Kaplan, E. D. Yorke, and J. A. Yorke, *J. Diff. Eq.* **49**:185–207 (1983).
8. P. Grassberger and I. Procaccia, *Phys. D* **9**:189–208 (1983).
9. H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro, *Phys. Rev. A(3)* **25**:3453–3456 (1982).
10. J. Guckenheimer, *Comm. Math. Phys.* **70**:133–160 (1979).
11. I. Gumowski and Ch. Mira, *Recurrences and Discrete Dynamic Systems*, Lecture Notes in Mathematics No. 809 (Springer-Verlag, Berlin, 1980).
12. G. Györgyi and P. Szépfalusy, *J. Stat. Phys.* **34**:451–475 (1984).
13. G. Györgyi and P. Szépfalusy, *Z. Physik B* **55**:179–186 (1984).
14. E. J. Hannan, *Time Series Analysis* (Methuen & Co., London, 1960), chap. II.3.
15. F. Hofbauer and G. Keller, *Math. Zeitschrift* **180**:119–140 (1982).
16. F. Hofbauer and G. Keller, *Ergod. Th. Dyn. Syst.* **2**:23–43 (1982).
17. I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables* (Wolters-Nordhoff Publishers, Groningen (NL), 1971).

18. J. L. Kaplan and J. A. Yorke, *Functional Differential Equations and Approximation of Fixed Points*, H. O. Peitgen and H. O. Walther, eds., *Springer Lect. Note. Math.* **730**:223–237 (1979).
19. G. Keller, Spectral analysis of fully developed chaotic maps, unpublished notes (1984).
20. G. Keller, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **69**:461–478 (1985).
21. T. Y. Li and J. A. Yorke, *Amer. Math. Month.* **82**:985–992 (1975).
22. E. N. Lorenz, *J. Atmos. Sci.* **20**:130–141 (1963).
23. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
24. R. M. May, *Science* **186**:645–647 (1974).
25. R. M. May, *Nature* **261**:459–467 (1976).
26. D. Mayer, *Comm. Math. Phys.* **95**(1):1–15 (1984).
27. D. Mayer and G. Roepstorff, *J. Stat. Phys.* **31**:309–326 (1983).
28. M. Misiurewicz, *Publ. Math. IHES* **53**:17–51 (1981).
29. S. E. Newhouse, *Entropy and Smooth Dynamics*, Lecture Notes in Physics No. 179 (Springer-Verlag, Berlin, 1983), 165–179.
30. H. E. Nusse, Chaos, yet no chance to get lost. Thesis, University of Utrecht (NL) (1983).
31. J. C. Oxtoby, *Measure and Category*, Graduate Texts in Mathematics, Vol. 3 (Springer-Verlag, Berlin, 1971).
32. W. Parry, *Topics in Ergodic Theory* (Cambridge University Press, Boston, 1981).
33. K. Petersen, *Ergodic Theory* (Cambridge University Press, Boston, 1983).
34. W. Philipp and W. Stout, *Mem. Am. Math. Soc.* **161** (1975).
35. R. H. Randles and D. A. Wolfe, *Introduction to the Theory of Nonparametric Statistics* (Wiley, New York, 1979).
36. C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors* in Applied Mathematical Sciences Vol. 41 (Springer-Verlag, Berlin, 1982).
37. W. Szlenk, I, *Bol. Soc. Math. Mexico* **24**:57–82 (1979); II, Ergodic Theory and Dynamical Systems II (College Park, Maryland, 1979/1980), p. 75–95; *Prog. Math.* **21** (1982).
38. V. A. Volkonski and Y. A. Rozanov, *Theory Prob. Appl.* **6**:186–198 (1961).
39. P. Walters, *An Introduction to Ergodic Theory* (Springer-Verlag, Berlin, 1982).
40. G. B. Wetherill, *Intermediate Statistical Methods* (Chapman and Hall, London/New York, 1981).
41. K. Yoshihara, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **35**:237–252 (1976).